

From digital library to institutional repository: a brief look at one library's path

Marianne A. Buehler
Marcia S. Trauernicht

Marianne A. Buehler is the Head of Publishing and Scholarship Support Services, Wallace Library, at Rochester Institute of Technology (RIT). She is a DSpace administrator for the RIT Digital Media Library (DML), also responsible for garnering content and managing the DML's documentation and structure.

Marcia S. Trauernicht is the Head of Cataloging and Special Assistant to the Director, Wallace Library, at the Rochester Institute of Technology (RIT.) She is a member of the RIT Digital Media Library (DML) team and is responsible for the standardization of metadata content and structure of the DML and many other digital collections supported by the library.

Abstract

Purpose – The purpose of this paper is to give an overview of the initial transformation of one academic library's digital media library into the institutional repository (IR) of its entire academic organization.

Design/methodology/approach – Description of an academic digital library's evolution into an institutional repository, accompanied by both practical and philosophical analyses.

Findings – Materials supporting an academic institution's scholarship (from colleges and from supporting organizations on campus) can play an important part in the success of an institutional repository. Standards for metadata content, data structures and hierarchies of collections must be followed consistently, and adequate storage for digital media within an institutional repository is necessary to have in advance of anticipated demand for the smooth operation of the IR and continued access to materials. Gray areas between the role of an institutional repository and the role of that organization's archives need to be addressed in the future.

Practical implications – Academic libraries can be highly successful in producing an institute repository by developing relationships with various organizations on campus in addition to the academic programs. Maintaining standards throughout the IR is crucial to future growth in an organized and consistent manner. Philosophical considerations of the role of the IR should be addressed in the beginning stages of the development of the IR for eliminating confusion and duplication of its contents with other campus organizations.

Originality/value – This paper describes the development of a digital library, created and maintained by an academic library, and its gradual change into a de facto institutional repository. Other libraries in the planning stages or initial steps of creating a campus-wide digital library or an institutional repository can benefit from the description of possible successes and problems that they could encounter during implementation.

Keywords Institutional repositories, Digital libraries, Archives

Paper type Case study

Garnering content for an institutional repository (IR) encompasses many detailed aspects ranging from engaging university producers of scholarly content through the submission process. Each institution has to make its own philosophical decisions of what items and types of materials are to be included in an IR, which then precipitates the need for other determinations that will fulfill the mission of an IR. In addition to faculty e-prints and student electronic theses and dissertations (ETDs), the Rochester Institute of Technology (RIT) Libraries have encouraged tangential scholarly IR items such as open access journals (RIT), academic newsletters, campus-wide publications, and other academic or administrative documents to be included. Each community, such as a college or a campus division, is encouraged to determine what item-types it wants to include in its space. This does not detract from the scholarly nature of the repository for the academic side.

Another question to contemplate is whether or not historical documents that already exist in the RIT Archive should be also made available online in the IR. Standardizations of hierarchies, metadata, and taking into consideration the unique needs of each community are important advisements to garner a critical mass of materials' support from college and non-college divisions.

Institutional repositories are important for the managing and dissemination of a university's intellectual property as part of its information assets strategy. Digital repositories typically preserve and showcase an institution's research (faculty and student), presentations, images, teaching materials, and administrative documents. These items are searchable and retrievable, deriving maximum benefit from the repository. The open access standard enhances these opportunities for the use of archived research, increases a global likelihood of collaborations among different disciplines, and provides for potential learning experiences (Hayes, 2005). The RIT Libraries define their open access institutional repository in a broad sense as an archive that houses peer-reviewed intellectual work, related to research and teaching that also includes materials representing RIT's cultural, historical, and administrative documentation. The level of peer-review is based on a continuum starting at a basic level of faculty-to-faculty or faculty-to-student to that of the classic journal editor review.

Given the positive attributes of a university IR, the RIT Libraries (made up of Wallace Library, the Melbert B. Cary, Jr. Graphic Arts Collection and the RIT Archive) embarked upon creating and establishing an IR (the RIT Digital Media Library (RIT DML)) in late 2002. Over 40% of the library staff was involved in collaborative teamwork that resulted in initially surveying RIT faculty and staff for their interest and archiving needs, then researching IR platforms (DSpace was selected) and various item formats, creating a custom "look," and garnering buy-in from RIT constituencies (see Figure 1).

Take in Figure (No.1)

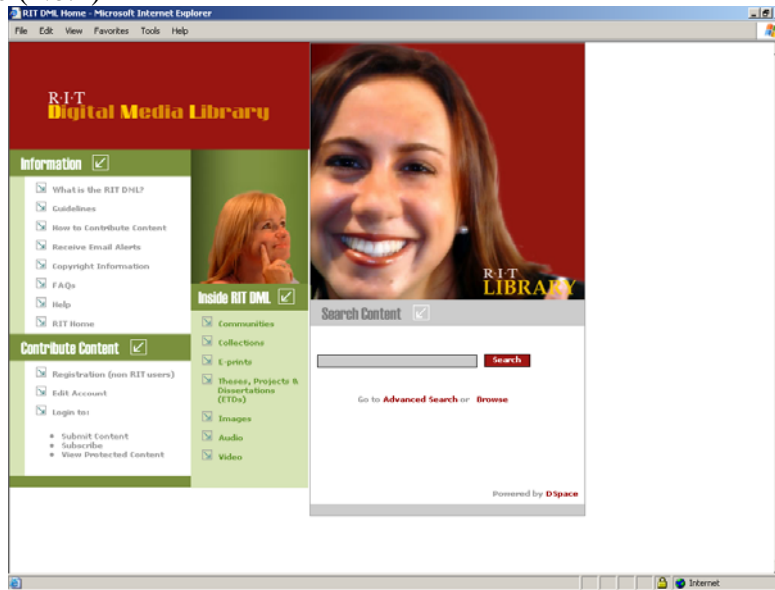


Figure 1.

Customized DSpace software for a personalized and colorful “look”

One of the best advantages for the Libraries’ creation of an IR at that point in time was that the RIT administration was simultaneously developing its strategic plan that included increasing RIT faculty scholarship and increasing research among faculty and with students. Dissemination of the scholarship was also an imperative (Buehler, 2005).

Reflecting upon the previous four years’ evolution of the RIT DML’s HYPERLINK "http://ritdml.rit.edu" conception, the RIT Libraries’ IR philosophical views, processes and activities have evolved to include additional types of materials, which in turn, have forged new and positive relationships with RIT’s academic departments, the university’s eight colleges, and the Libraries. Many faculty members have research, articles,

conference proceedings, monographs, and reports with citations only listed on their department or personal web pages. Few have full-text available. The upkeep of websites can be a time and priority challenge. The RIT Libraries' offer to post and maintain the full text of scholarly, peer-reviewed materials will continue to prove a worthwhile opportunity to many of the faculty. In this article, the RIT DML and RIT's IR are referred to as one in the same.

After the creation of the RIT DML site that mirrors the structure of RIT (each of the eight colleges have their own community, as does academic support departments, such as the RIT Libraries, Academic Affairs, Student Affairs, Office of the President, RIT Governance, University News, and others), the next step was to commence creating hierarchies that would reflect how each college or division was set up. A hierarchy was based overall (in seven of the colleges) on the particular college and its departments, breaking the departmental sub-communities into faculty scholarship, student scholarship, and administrative document collections. By following the departmental hierarchy of each college and the format of separating scholarship by category of author (faculty, student and administration) within the collections, the library staff endeavored to create a community and collection structure that could be retained as new programs of study are added in the future, and as current programs change or merge with other programs within the college (see Figure 2).

Take in Figure (No. 2)

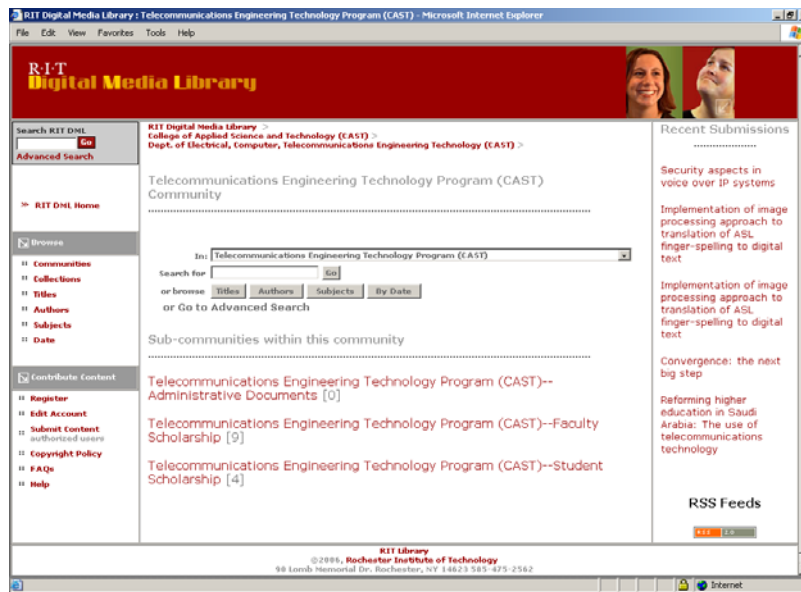


Figure 2.

Delineating types of materials within each sub-community for consistent access

This approach makes it much simpler for the library staff to update each college community without the need to make multiple changes throughout the IR, saving staff time, and the standardized structure provides consistency and reliability of access for those searching within the IR. The eighth college preferred to group its scholarship into disciplines or areas of research, with sub-topics under each. The reasoning given behind this departure of the RIT IR's college community structure or hierarchy was to emphasize scholarship that cut across departments and units of that college rather than separate faculty and staff scholarship on the same topic by a purely administrative structure. Since so many faculty and staff members of that college collaborate across departments on a regular basis, it was felt that those searching for its scholarship by entering its repository community would not be able to locate the materials in a specific area of research

through a departmental hierarchy. After much deliberation on this request, the library staff set up the eighth college's hierarchy as requested. A year after this structure was implemented, requests to change the subject hierarchy to that of departmental/administrative, identical to the other colleges, came pouring in from both the college and library staff. The common complaint made was that no one could find anything. This episode reinforced the need to retain standardization of hierarchies within the IR for the same types of communities, and that the library staff should not depart from its role as the administrator of the IR in maintaining that standardization.

As has been noted in many publications, at conferences, and within collegial conversations, it is common knowledge that garnering faculty and research content is a challenge. At RIT, the colleges span subject/research areas that range from the sciences (chemistry, physics, etc.), engineering/engineering technology, business, computer science and information technology to liberal arts, hospitality/tourism, printing, ceramics, photography, and furniture design to Deaf culture and American Sign Language. Given the wide variety of types of scholarship that are engendered, the variables rise on why some faculty members are willing to have their work posted and others are not. Some of the more notable and universally experienced basis RIT faculty give as reasons of resistance to using an IR are: the time and priority factor, a lack of awareness of or unconvinced of the immediate value, copyright retention concerns, and the potential impediment or undermining of the traditional scholarly publishing system that faculty

rely on for tenure and professional recognition, as faculty members also prefer to exert full control over their scholarly content (Cervone, 2004).

To alleviate as many of the faculty IR buy-in concerns as possible, the Publishing and Scholarship Support Services department, as one of its main missions that focuses on supporting scholarly activities (that includes offering dissemination venues) is to manage various aspects of the IR, including assisting faculty submit their scholarship in the IR. Students who work in the actual physical space of the Publishing and Scholarship Support Center (PSSC) HYPERLINK "<http://wally.rit.edu/userservices/pubschol/>" post the vast majority of RIT materials in the repository (see Figure 3).

Take in Figure (No.3)

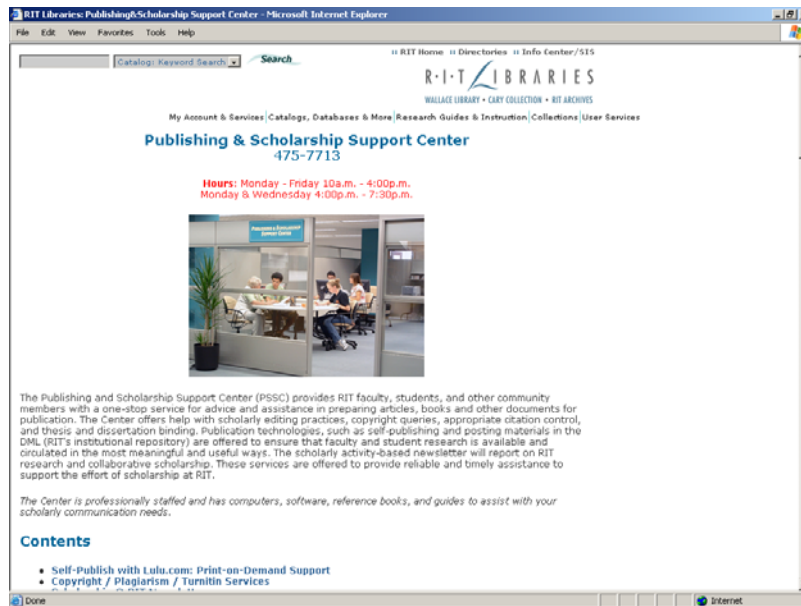


Figure 3.

The Publishing and Scholarship Support Center

The College of Science (COS) Dean has given permission to harvest and post all COS faculty scholarship. Since the faculty members have already posted or published their materials in other venues, PSSC staff members are harvesting faculty work wherever feasible, taking into consideration copyright ownership. PSSC student employees harvest metadata and documents in Cornell University's open access arXiv that maintains e-prints in physics, mathematics, computer science and quantitative biology, and then post the COS faculty pre-prints, postprints and reports in the IR. If the documents have been also published in a journal, that information is included in the metadata.

In instances where RIT faculty members have published in journals that require the author's post-print (which seems to be most common), the PSSC staff requests the article from the faculty. Unfortunately, many authors do not retain their final copy, especially for retrospective materials. When the final e-print is not available, the PSSC posts the metadata only. When the journal necessitates its final post-print in an IR (less common), that copy is posted and additional, required publication documentation, such as journal copyright information and or a link to the online journal is included. An option for e-prints in an instance where the author does not own the copyright, is to input the metadata and include a PDF of the journal's policy that precludes open access in an IR, so that users are aware of why the full-text is absent. PSSC student employees are trained to use the University of Nottingham's SHERPA RoMEO website [HYPERLINK "http://www.sherpa.ac.uk/romeo.php"](http://www.sherpa.ac.uk/romeo.php) to assist in determining journal/publisher

qualifications for publisher copyright self-archiving policies. They then PDF a particular journal policy web page and post it in place of the full-text document (see Figure 4).

Take in Figure (No. 4)



Figure 4.

This document replaces the full-text when journal copyright precludes IR posting.

In other cases, faculty article metadata is input with a direct link to an article, such as what is required from the American Chemical Society's website publications.

HYPERLINK "<http://acswebcontent.acs.org/home.html>" The Libraries as publisher

assure basic standards by ensuring copyright compliance and by including sufficient metadata and the full-text where possible, in order to provide optimum use.

Over the past few years, the academic community has observed publishers, such as Elsevier, who have changed their open access policy to allow authors to post their personal scholarship on an institutional server. Given the evolving nature of open access and publishers' policies, the philosophy of the Head of the PSSC is to post the metadata and the Sherpa documentation (as the uploaded file for an item). As each publisher recognizes the open access movement for the greater good of globally sharing scholarly work by permitting posting in an IR, the Sherpa document can be replaced with an author's final work.

The PSSC also accepts electronic or paper materials from faculty. In addition to preprint and post-print articles, monographs, book chapters, reports, poetry, workbooks, conference proceedings (including PowerPoint presentations), symposiums, websites, faculty-generated newsletters, bibliographies, RIT faculty theses and dissertations, and faculty-recommended exemplary undergraduate works are posted.

The RIT Libraries recently worked with the College of Applied Science and Technology in launching the Libraries' first open access publication, *The Journal of Applied Science and Engineering Technology* (JASET). In using the Open Journal Software (OJS), for journal-creation, one of the requirements is to post all articles in the university's IR. In spring 2007, the first issue of JASET will be posted in RIT's IR (see Figure 5).

Take in Figure (No.5)

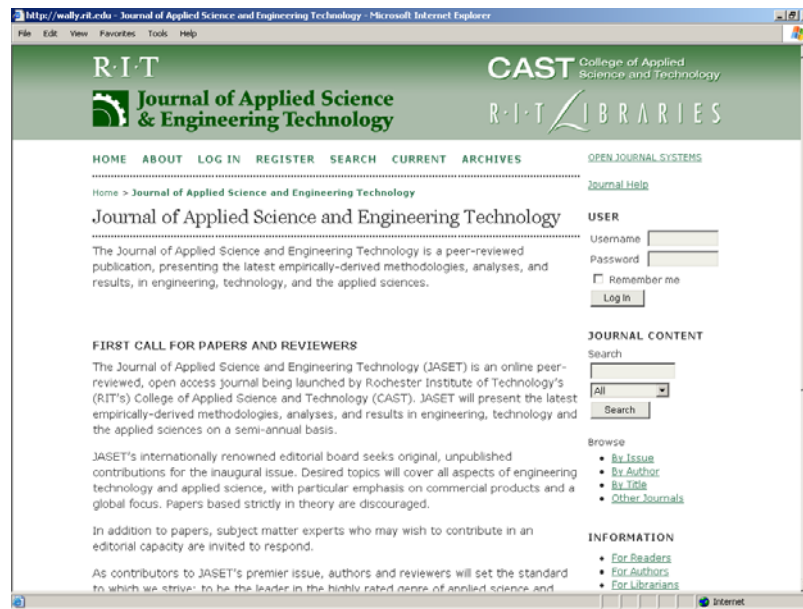


Figure 5.

JASET will be posted in the RIT DML in spring 2007

Student theses, dissertations, and capstone projects (or Electronic Theses and Dissertations (ETDs)) are far less complicated to post, as a student owns his or her paper's copyright and tends to see an immediate value for using the IR as a current or future career tool. The PSSC staff posts the ETDs into the RIT DML as an expansion of its service of binding theses and dissertations for RIT students. At this time, mandatory electronic submission is not a requirement for the completion of a master's or doctorate degree at RIT, and is purely voluntary. (A print copy is accepted as the document of record at this time, and retained in the RIT Archive.) RIT graduates, in growing numbers, have been very eager to have their scholarship posted at a permanent and freely-accessible site so that the URIs (uniform resource identifiers) of those works can be given to potential employers as part of the graduates' portfolios of work. From September 2005

to September 2006, 58% (160) of the 275 bound theses and dissertations received within that time period were submitted to the IR. Electronic access to these works provides immediate and potentially world-wide recognition for the authors, and savvy students have taken advantage of the opportunities provided to them by the DML. Currently, digital images that may accompany an ETD are available in the RIT DML, but the storing of and access to audio, video and multimedia files have yet to be refined and substantially supported (see Figure 6)

Take in Figure (No.6)

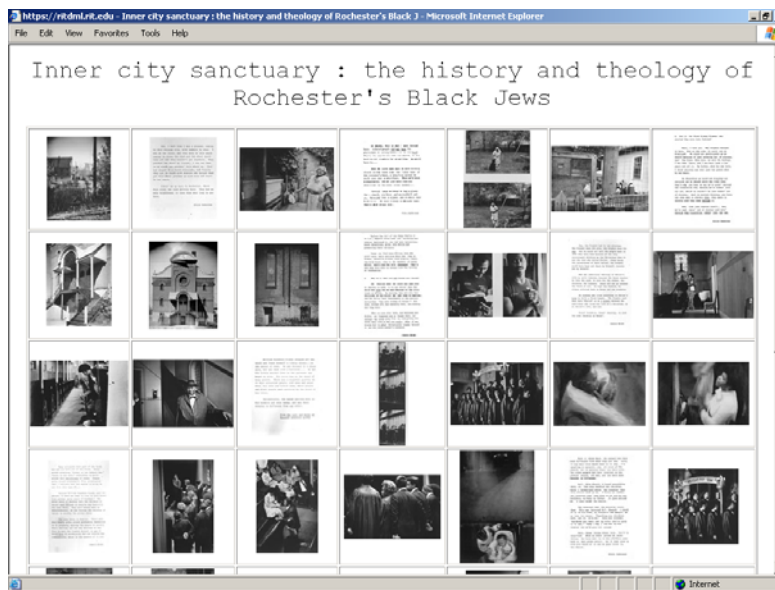


Figure 6.

Accompanying digital photos to a 1978 RIT thesis

Executable files, primarily in the area of Computer Science, also accompany a number of works. The problems of supporting such files plus security issues continue to be studied.

Coordination of the RIT DML, the processing of ETDs and submission of those works to the UMI/ProQuest dissertation services are now being addressed by the library staff.

While the PSSC has posted many faculty documents and graduate student work in the RIT DML, there are other divisions on campus that have scholarly-related materials that either report on scholarly activities, support aspects of scholarly work within its content, or are creative works in themselves. Because of the scholarly nature of the content of these publications, it was decided to include these peripheral and integral items whether they are born digital or in print. In some cases, digital files are received and in others, the actual print documents are scanned. The non-college divisions have been surprisingly supportive and interested in participating, with one staff member stating, “We want to be part of the RIT DML family.” The student magazine, *The Reporter*, has issues dating back to 1951. From 1951-1963, the publication was in a large 12”x17” format. The Publishing and Scholarship Support Services department purchased a large-bed model scanner to accommodate the 3500+ issues that will be posted in the IR. *The Reporter*’s editors have been concerned about the lack of a sustainable archival system for their publication. They are pleased that RIT’s IR will maintain the magazine in perpetuity and will also serve the needs of the many people who contact them in order to locate previously published articles (see Figure 7).

Take in Figure (No.7)



Figure 7.

A 1951 digitized copy of the *Reporter* student magazine

The University News community was the first to request having its biweekly newsletter (*RIT news&events*) posted in the DML. News items typically published focus on faculty scholarship that includes publications, upcoming or prior presentations, lectures/poetry readings, viewpoints, reports on current research, academic contests/awards, and other newsmaker information. To date, the Publishing Center's assistant has input files from October 2002-December 2004. Keywords are alphabetized for easy review for those preferring to browse a citation.

Other publications that are IR-worthy and have been incorporated in the DML are the

Academic Support Center Quarterly that focuses on topics such as teaching and learning, academic assessment, learning disabilities, and student success. *Signatures*, RIT's journal of art and literature, encompasses creative poetry, images, photography, and more by students. In recent years, *Signatures'* web version has a multimedia component that the PSSC hopes to capture in the DML. The PSSC students also post the *University Magazine*, the official publication for RIT alumni, issued fall, winter and spring. It promotes an alumni connection with its current academic research, cultural activities, ongoing alumni services, and on individual alumni successes. The *Scholarship@RIT* newsletter, a PSSC publication that highlights the Libraries' *Faculty Scholar Series'* presentations, research, teaching projects, faculty sabbaticals, showcases student work, and provides copyright and licensing articles, is also posted in the repository (see Figure 8).

Take in Figure (No.8)

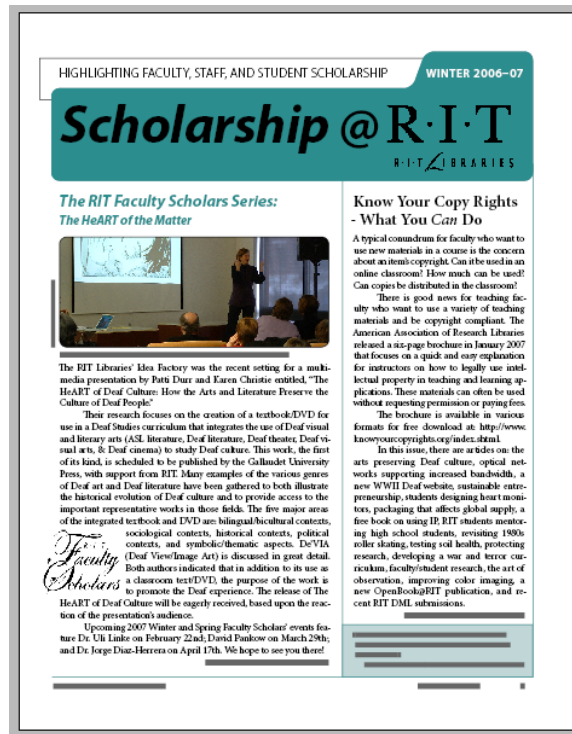


Figure 8.

The *Scholarship@RIT* newsletter is available in RIT's institutional repository

Test Targets magazine and *The Scholarship Report* are valuable faculty-created contributions to the IR. The most recent addition to the DML is the *ITS* (Information Technology Services) *News*, that details current IT projects at RIT and how to use various products and services supported by ITS.

Aside from the easy availability of RIT-created documents in the IR, one of the most useful aspects of providing open access to the publications is the built-in DSpace PDF search tool that allows text that has been PDFed with optical character recognition (OCR) software to search for keywords in all items. It makes for an effective search that returns

all of the items that have the requested key words. Using the search function in a PDF document (binoculars icon), one can then search an individual item using the keywords, finding each instance of its use.

Having an efficient search engine to use when searching within archived administrative documents is an advantage when trying to locating topics that are potentially spread out among many collections of documents. RIT's Academic Senate (AS) worked with the Publishing and Scholarship Support Services to determine a hierarchical system that would serve to provide a transparency of seeking AS documentation to the RIT community, avoiding the pitfalls of a non-searchable website interface. The AS secretary now emails links to the RIT staff that point to the document citation in the IR instead of sending attached documents that take up valuable space in individual email accounts. There is potential for RIT's 900 faculty members to access the IR on a regular basis to read the AS documents and become interested in the IR's concept for their own purposes. For AS working documents that are restricted to AS Senators-only, the top level IR administrator set up an LDAP-authenticated collection for their use. Other RIT governance bodies are currently exploring the possibilities of also posting administrative materials in the DML, as are some college divisions. The plan is to also "lock down" these other collections for RIT use only. The citations are viewable. When a user clicks on the View/Open link to read the document, a screen pops up asking for the person to login (see Figure 9).

Take in Figure (No. 9)

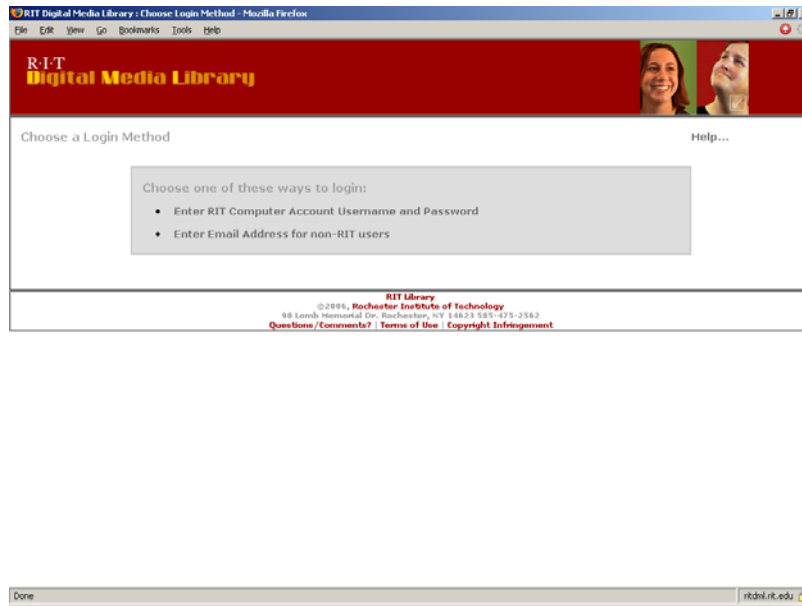


Figure 9.

The RIT DML authentication screen asking users to login

As the RIT DML has grown and developed over the past four years into an institutional repository, an overlapping of its scope with that of the RIT Archive has become very apparent. As stated earlier, the RIT DML has begun acquiring peripheral materials that support RIT scholarship. Most of these materials are also historical in nature, and many are housed in the RIT Archive in their original print versions. And, all of RIT's master's theses and Ph.D. dissertations are also maintained there, again in hard copy.

Concern on both the RIT DML staff's and the RIT Archivist's parts has brought about many ongoing discussions on what types of historical institute materials should be

present in the IR or are more appropriately supported by a separate Archives records management system or via the Archive's own website. Briefly, a records management system provides "systematic and administration control of records throughout their life-cycle to ensure efficiency and economy in their creation, use, handling, control, maintenance, and disposition." (Pearce-Moses 2005) Selected materials or records are archived because of their historical value to an organization. Both the RIT Archive and the RIT DML agree on providing permanent access to RIT-related materials, but concern over duplication of effort and appropriateness of some materials (though not because of sensitive information) being present within the RIT DML versus the appeal of easy accessibility and a highly visible presence of archival materials via the RIT DML has not been resolved on the RIT Libraries' part. At this time, neither the RIT Archive nor RIT itself has a fully developed records management system, so all have concurred to post some materials as the best approach for both the RIT DML and the RIT Archive at this time. As an RIT Archive records management system evolves, this discussion will resume.

The inclusion of all IR items require workable hierarchies and appropriate metadata schema. Within the RIT DML, the area of metadata continues to evolve and expand, as more varied types of materials are added to the IR. For each new format of material added to the RIT DML, the Head of Cataloging establishes additional enhanced Dublin Core fields appropriate to that material, and those fields are added to the metadata

template for that format by the Library's programmer. For eprints and ETDs, the author's keywords are utilized whenever they are supplied. ETDs also receive Library of Congress subject headings, since it is a current RIT requirement that all graduate theses and dissertations be fully cataloged in the online public access catalog (OPAC) (see Figure 10)

Take in Figure (No.10).

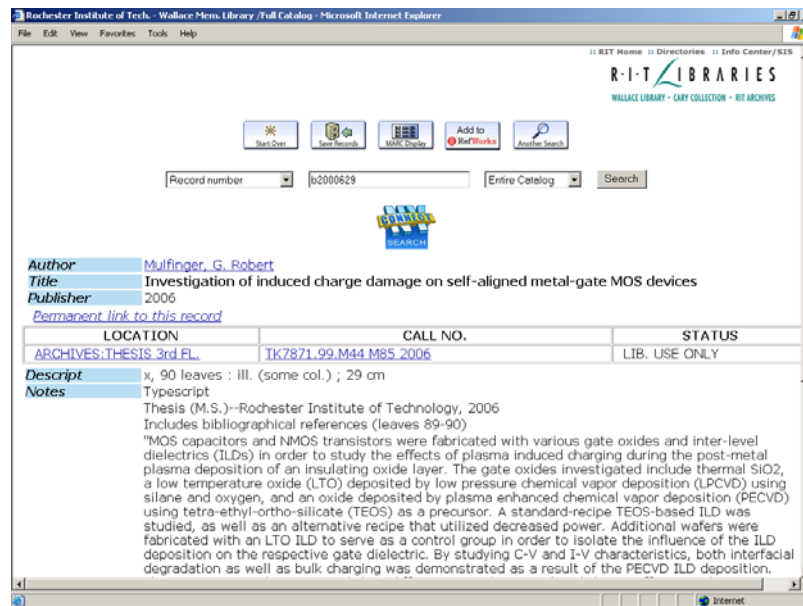


Figure 10.

OPAC view of ETD that has been cataloged

RIT's Head of Cataloging establishes metadata guidelines for each new type of material to be added to the IR, and often creates one or two prototype records as templates for the PSSC staff members to emulate as they post recurrent items such as newsletter issues, etc. PSSC staff often supplies additional keywords for these materials to enhance

keyword access, even though each individual issue can be searched full-text. A future goal is to utilize discipline-specific established thesauri terms for the scholarly work in various disciplines in addition to maintaining keyword access, although the demand from the colleges has been very light at most for this type of enhancement.

Server space set aside for the RIT DML has been outgrown since the IR's inception.

Within the past year, items had to be held for a time as the Library acquired additional server space. Text-only items, images and just one or two small video clips have been accepted, since the RIT DML could not accommodate large video or audio files in an adequate manner. Currently work is being done on a project that will allow access from the DML to large video files housed on a separate server. In order to serve the immediate needs of the campus community, this approach may be utilized on a larger scale.

However, the long-term security of archiving institute repository files in a separate location that is not formally associated with the IR must be addressed.

By archiving and preserving faculty, staff, and student peer-reviewed scholarly materials, the RIT Libraries are providing value for the creative output of RIT scholarship on a personal, departmental, college, and university level. The goal is to create a comprehensive, well designed archive with useful metadata to showcase RIT's intellectual, historical, and cultural materials, in a variety of formats. By garnering content from all areas of the campus to highlight the university's creative output in ways

that encourage collaboration, exposure, and value within RIT and on a national and global level, the RIT Libraries will have achieved their goal of an institutional repository.

References

- Buehler, M.A. & Boateng, A. 2005, "The evolving impact of institutional repositories on reference librarians," *Reference Services Review*, vol. 33, no. 3, pp. 291.
- Cervone, H.F. June 1, 2004, "The Repository Adventure," *Library Journal*.
- Hayes, H. August 2005, *Digital Repositories, Helping universities and colleges*, JISC, Scotland.
- Pearce-Moses, R. 2005, *A Glossary of Archival and Records Terminology*. The Society of American Archivists.